

# RATES IN ALMOST SURE INVARIANCE PRINCIPLE FOR YOUNG TOWERS WITH EXPONENTIAL TAILS

ALEXEY KOREPANOV

**ABSTRACT.** We prove the almost sure invariance principle with rate  $o(n^{1/p})$  for every  $p$  for Hölder continuous observables on nonuniformly expanding and nonuniformly hyperbolic transformations with exponential tails. Examples include logistic maps at Collet-Eckmann parameters and dispersing billiards.

As a part of our method, we build Hölder continuous semiconjugacies between nonuniformly expanding transformations and Young towers where the base transformation is a Bernoulli shift. For this we do not require exponential tails.

## 1. INTRODUCTION

**Definition 1.1.** We say that a random process  $X_0, X_1, \dots$  satisfies the *Almost Sure Invariance Principle* (ASIP) with rate  $o(n^\beta)$ ,  $0 \leq \beta < 1/2$ , if without changing the distribution,  $\{X_n, n \geq 0\}$  can be redefined on a new probability space with a Brownian motion  $B_t$ , such that

$$X_n = B_n + o(n^\beta) \quad \text{almost surely.}$$

The ASIP is a strong statistical property. It implies the functional central limit theorem, the functional law of iterated logarithm and other statistical laws (see e.g. [15, Chapter 1]).

Rates in the ASIP reflect the strength of the approximation, and they received a significant amount of attention, see [1] for the history and references. In dynamical systems, the ASIP has been shown to hold in various settings [3, 5, 6, 7, 8, 13, 14].

In this paper we work with nonuniformly expanding transformations  $T: \Lambda \rightarrow \Lambda$  as in [17], such as logistic maps with Collet-Eckmann parameters, Gibbs-Markov maps with big images and intermittent maps.

Suppose that  $\nu$  is the unique  $T$ -invariant ergodic physical measure,  $v: \Lambda \rightarrow \mathbb{R}$  is a Hölder continuous observable with  $\int v d\nu = 0$ , and  $v_n = \sum_{k=0}^{n-1} v \circ T^k$ . Then  $v_n, n \geq 0$  is a stationary random process on the probability space  $(\Lambda, \nu)$ .

We prove that if  $T$  has *exponential tails*, then  $v_n$  satisfies the ASIP with rate  $o(n^{1/p})$  for every  $p > 2$ . This result carries over to nonuniformly hyperbolic systems with exponential tails and uniform contraction along stable leaves as in Young [16], such as dispersing billiards.

In the class of nonuniformly expanding and nonuniformly hyperbolic systems with Hölder continuous observables we improve the best previously available rate  $O(n^{1/4}(\log n)^{1/2}(\log \log n)^{1/4})$  due to Cuny and Merlevède [4].

Our strategy is to represent  $T: \Lambda \rightarrow \Lambda$  as a Bernoulli shift via a semiconjugacy, and use Komlós-Major-Tusnády approximation under dependence by Berkes, Liu and Wu [1].

*Remark 1.2.* Chernov and Haskell [2] prove Bernoulli property for K-mixing nonuniformly hyperbolic maps. That is, such maps are measure-theoretically isomorphic to Bernoulli shifts. They remark that even though the Bernoulli property is a characterization of

extreme chaotic behaviour, it is not helpful in proving statistical properties such as the central limit theorem. This is because a measure-theoretic isomorphism alone does not have to preserve any useful information about the structure of the space, such as metric or coordinates.

In contrast, we build a semiconjugacy to a Bernoulli shift which preserves enough information to prove the ASIP.

As a part of our method, we show that nonuniformly expanding transformations can be modelled by Young towers, where the base map is a Bernoulli shift. Therefore they can be studied as (Harris recurrent) Markov chains. For this we do not require exponential tails.

## 2. STATEMENT OF THE RESULT

We use notation  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}_0 = \{0, 1, \dots\}$ . All functions, subsets and partitions are assumed to be measurable. When we work with metric spaces, the default sigma algebra is Borel, and for finite and countable spaces the sigma algebra is discrete.

Let  $(\Lambda, d_\Lambda)$  be a bounded complete metric space and  $T: \Lambda \rightarrow \Lambda$ . Let  $Y$  be a closed subset of  $\Lambda$  and  $m$  be a probability measure on  $Y$ . Let  $\alpha$  be an at most countable partition of  $Y$  (modulo a zero measure set) such that  $m(a) > 0$  and  $m(\bar{a} \setminus a) = 0$  for all  $a \in \alpha$ , where  $\bar{a}$  is the closure of  $a$ .

Let  $\tau: Y \rightarrow \mathbb{N}$  be an integrable function which is constant on each  $a \in \alpha$  with value  $\tau(a)$  such that  $T^{\tau(y)}(y) \in Y$  for every  $y \in Y$ . We denote  $T_Y: Y \rightarrow Y$ ,  $T_Y(y) = T^{\tau(y)}(y)$ .

We assume that there exist constants  $0 < \eta \leq 1$ ,  $\lambda > 1$  and  $C_\ell, K \geq 1$ , and for each  $a \in \alpha$  there exists a map  $T_a: \Lambda \rightarrow \Lambda$  such that  $T_a^\ell = T^\ell$  for all  $0 \leq \ell < \tau(a)$   $m$ -almost surely on  $\bar{a}$ , and  $T_{Y,a} = T_a^{\tau(a)}$  is a bijection between  $\bar{a}$  and  $Y$ . Also, for all  $x, y \in \bar{a}$ :

- $d_\Lambda(T_{Y,a}(x), T_{Y,a}(y)) \geq \lambda d(x, y)$ ,
- $d_\Lambda(T_a^\ell(x), T_a^\ell(y)) \leq C_\ell d_\Lambda(T_{Y,a}(x), T_{Y,a}(y))$  for all  $0 \leq \ell < \tau(a)$ ,
- $T_{Y,a}: \bar{a} \rightarrow Y$  is nonsingular and its inverse Jacobian  $\zeta = \frac{dm}{dm \circ T_{Y,a}}$  satisfies

$$|\log \zeta(x) - \log \zeta(y)| \leq K d_\Lambda^\eta(T_{Y,a}(x), T_{Y,a}(y)).$$

We say that  $T: \Lambda \rightarrow \Lambda$  is a *nonuniformly expanding* map. Our definition is more restrictive than in [12]: we are using additional structure such as completeness of  $\Lambda$  to ensure Lipschitz semiconjugacy with suspensions over Bernoulli shifts.

It is standard [17] that there is a unique  $T$ -invariant ergodic probability measure on  $\Lambda$ , with respect to which  $m$  is absolutely continuous. We denote this measure by  $\nu$ .

For an observable  $v: \Lambda \rightarrow \mathbb{R}$ , denote

$$|v|_\infty = \sup_{x \in \Lambda} |v(x)|, \quad |v|_\eta = \sup_{x \neq y \in \Lambda} \frac{|v(x) - v(y)|}{d^\eta(x, y)} \quad \text{and} \quad \|v\|_\eta = |v|_\infty + |v|_\eta.$$

We say that  $v$  is *centered*, if  $\int v d\nu = 0$ , and that  $v$  is Hölder, if  $\|v\|_\eta < \infty$ .

Our main result is:

**Theorem 2.1.** *Suppose that there exists  $\beta > 0$  such that  $\int_Y e^{\beta\tau} dm < \infty$ . If  $v: \Lambda \rightarrow \mathbb{R}$  is a Hölder centered observable, then the process  $v_n = \sum_{k=0}^{n-1} v \circ T^k$ , defined on the probability space  $(\Lambda, \nu)$ , satisfies the ASIP with rate  $o(n^{1/p})$  for every  $p > 2$ .*

*Remark 2.2.* In nonuniformly hyperbolic transformations with exponential tails and uniform contraction along stable leaves as in Young [16], Hölder observables reduce to Hölder observables on the nonuniformly expanding quotient system through a bounded coboundary. A detailed exposition can be found in [12, Section 5]. Thus Theorem 2.1 implies the

ASIP with rate  $o(n^{1/p})$  for every  $p$  for maps in [16], such as dispersing billiards or Hénon maps.

The paper is organized as follows: in Section 3 we introduce the notion of *Bernoulli Young towers* and state Theorem 3.4 which establishes a semiconjugacy between  $T: \Lambda \rightarrow \Lambda$  and a Bernoulli Young tower. Theorem 3.4 is proved in Section 4. We prove Theorem 2.1 in Section 5.

### 3. BERNOULLI YOUNG TOWERS

Suppose that

- $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$  is a finite or countable probability space,
- $h_{\mathcal{A}}: \mathcal{A} \rightarrow \mathbb{N}$  is an integrable function,
- $0 < \xi < 1$  is a constant.

Define a probability space  $(X, \mathbb{P}_X) = (\mathcal{A}^{\mathbb{N}}, \mathbb{P}_{\mathcal{A}}^{\mathbb{N}})$  and let  $f_X: X \rightarrow X$  be the left shift,

$$f_X(a_0, a_1, \dots) = (a_1, a_2, \dots).$$

Define  $h: X \rightarrow \mathbb{N}$ ,  $h(a_0, a_1, \dots) = h_{\mathcal{A}}(a_0)$ . Let  $f: \Delta \rightarrow \Delta$  be a suspension over  $f_X: X \rightarrow X$  with a roof function  $h$ , i.e.

$$(1) \quad \Delta = \{(x, \ell) \in X \times \mathbb{Z} : 0 \leq \ell < h(x)\}$$

$$f(x, \ell) = \begin{cases} (x, \ell + 1), & \ell < h(x) - 1, \\ (f_X(x), 0), & \ell = h(x) - 1. \end{cases}$$

Define a distance  $d$  on  $X$  by  $d(x, y) = \xi^{s(x, y)}$ , where  $s: X \times X \rightarrow \mathbb{N}_0$  is the separation time,

$$s((a_0, a_1, \dots), (b_0, b_1, \dots)) = \inf\{j \geq 0 : a_j \neq b_j\}.$$

Let  $d$  also denote the natural compatible distance on  $\Delta$ :

$$(2) \quad d((x, k), (y, j)) = \begin{cases} 1, & k \neq j \\ d(x, y), & k = j. \end{cases}$$

Let  $\bar{h} = \int h \, d\mathbb{P}$ . Let  $\mathbb{P}$  be the probability measure on  $\Delta$  given by  $\mathbb{P}(A \times \{\ell\}) = \bar{h}^{-1} m(A)$  for all  $\ell \geq 0$  and  $A \subset \{y \in Y : h(y) \geq \ell + 1\}$ . Note that  $\mathbb{P}$  is  $f$ -invariant.

Let  $\Delta_k = \{(y, \ell) \in \Delta : \ell = k\}$ . Then  $X$  is naturally identified with  $\Delta_0$ , which we refer to as the *base* of the suspension, and  $\mathbb{P}_X, f_X$  have their counterparts on  $\Delta_0$ , which we also denote  $\mathbb{P}_X, f_X$ .

**Definition 3.1.** We call the map  $f: \Delta \rightarrow \Delta$  as above a (non-invertible) *Bernoulli Young tower*.

*Remark 3.2.* To define a Bernoulli Young tower, we need an at most countable probability space  $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$ , an integrable function  $h_{\mathcal{A}}: \mathcal{A} \rightarrow \mathbb{N}$  and a constant  $0 < \xi < 1$ . Further we always use notation for Bernoulli Young towers as above, i.e. with the symbols  $f, \Delta, \mathcal{A}, \mathbb{P}_{\mathcal{A}}, X, \mathbb{P}_X, f_X, h, \mathbb{P}, d, \xi$ .

*Remark 3.3.* Bernoulli Young tower is a particular case of the classical Young tower, when the base transformation is a Bernoulli shift.

Our key technical result is:

**Theorem 3.4.** *Suppose that  $T: \Lambda \rightarrow \Lambda$  is a nonuniformly expanding map. Then there exists a Bernoulli Young tower  $f: \Delta \rightarrow \Delta$  and a map  $\pi: \Delta \rightarrow \Lambda$  such that*

- $\pi$  is Lipschitz:

$$d_\Lambda(\pi(x), \pi(y)) \leq C_\Lambda d(x, y) \quad \text{for all } x, y \in \Delta,$$

where  $C_\Lambda = \lambda C_\ell \text{diam } \Lambda$ ,

- $\pi$  is a semiconjugacy:  $\mathbb{P}$ -almost surely,  $T \circ \pi = \pi \circ f$ ,
- $\pi$  preserves measures:  $\pi_* \mathbb{P}_X = m$  and  $\pi_* \mathbb{P} = \nu$ .

In addition, moments of  $h$  are closely related to those of  $\tau$ :

- (Weak polynomial moments) If there exist  $C_\tau > 0$  and  $\beta > 1$  such that  $m(\tau \geq \ell) \leq C_\tau \ell^{-\beta}$  for all  $\ell \geq 1$ , then  $\mathbb{P}_X(h \geq \ell) \leq C \ell^{-\beta}$  for all  $\ell \geq 1$ , where the constant  $C$  continuously depends on  $C_\tau$ ,  $\beta$ ,  $\lambda$ ,  $K$  and  $\eta$ .
- (Strong polynomial moments) If there exist constants  $C_\tau > 0$  and  $\beta > 1$  such that  $\int \tau^\beta dm \leq C_\tau$ , then  $\int h^\beta d\mathbb{P}_X \leq C$ , where the constant  $C$  continuously depends on  $C_\tau$ ,  $\beta$ ,  $\lambda$ ,  $K$  and  $\eta$ .
- (Exponential and stretched exponential moments) If there exist constants  $C_\tau > 0$ ,  $\beta > 0$  and  $\gamma \in (0, 1]$  such that  $\int e^{\beta\tau^\gamma} dm \leq C_\tau$ , then  $\int e^{\beta'h^\gamma} d\mathbb{P}_X \leq C$ , where the constants  $\beta' \in (0, \beta]$  and  $C > 0$  depend continuously on  $C_\tau$ ,  $\beta$ ,  $\gamma$ ,  $\lambda$ ,  $K$  and  $\eta$ .
- (Exactly exponential moments) If  $\int e^{\beta\tau} dm < \infty$  for some  $\beta > 0$ , then  $f: \Delta \rightarrow \Delta$  can be constructed so that

$$\mathbb{P}_X(h = n) = \begin{cases} \theta(1 - \theta)^{-1}(1 - \theta)^{n/N}, & n \in \{N, 2N, 3N, \dots\} \\ 0, & \text{else} \end{cases}$$

with some  $0 < \theta < 1$  and  $N \geq 1$ .

For  $v: \Delta \rightarrow \mathbb{R}$  and  $\eta \in (0, 1]$ , denote

$$|v|_\infty = \sup_{x \in \Delta} |v(x)|, \quad |v|_\eta = \sup_{x \neq y \in \Delta} \frac{|v(x) - v(y)|}{d^\eta(x, y)} \quad \text{and} \quad \|v\|_\eta = |v|_\infty + |v|_\eta.$$

#### 4. PROOF OF THEOREM 3.4

**4.1. Construction of Bernoulli Young tower.** We define  $\mathcal{A}$  as the set of all finite words in the alphabet  $\alpha$  (not including the empty word). For  $w = a_0 \dots a_{n-1} \in \mathcal{A}$  we define

$$|w| = n \quad \text{and} \quad h_{\mathcal{A}}(w) = \tau(a_0) + \dots + \tau(a_{n-1}).$$

Let  $Y_w = (T_{Y, a_{n-1}} \circ \dots \circ T_{Y, a_0})^{-1}(Y)$  and

$$T_{Y, w}: Y_w \rightarrow Y, \quad f_w = T_{Y, a_{n-1}} \circ \dots \circ T_{Y, a_0}.$$

*Remark 4.1.*  $Y_w$  is closed,  $m(\bar{Y}_w \setminus Y_w) = 0$  and  $\text{diam } Y_w \leq \lambda^{-n}$ .

We use the measure  $\mathbb{P}_{\mathcal{A}}$  from the following

**Lemma 4.2.** *There exists a probability measure  $\mathbb{P}_{\mathcal{A}}$  on  $\mathcal{A}$  and a disintegration  $m = \sum_{w \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w) m_w$ , where  $m_w$  are probability measures on  $Y$ , such that for every  $w \in \mathcal{A}$ ,*

- $m_w$  is supported on  $Y_w$ ,
- $(T^{h_{\mathcal{A}}(w)})_* m_w = m$ .

In addition,

- If there exist  $C_\tau > 0$  and  $\beta > 1$  such that  $m(\tau \geq \ell) \leq C_\tau \ell^{-\beta}$  for all  $\ell \geq 1$ , then  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq \ell) \leq C \ell^{-\beta}$  for all  $\ell \geq 1$ , where the constant  $C$  continuously depends on  $C_\tau$ ,  $\beta$ ,  $\lambda$ ,  $K$  and  $\eta$ .
- If there exist constants  $C_\tau > 0$  and  $\beta > 1$  such that  $\int \tau^\beta dm \leq C_\tau$ , then  $\int h_{\mathcal{A}}^\beta d\mathbb{P}_{\mathcal{A}} \leq C$ , where the constant  $C$  continuously depends on  $C_\tau$ ,  $\beta$ ,  $\lambda$ ,  $K$  and  $\eta$ .

- If there exist constants  $C_\tau > 0$ ,  $\beta > 0$  and  $\gamma \in (0, 1]$  such that  $\int e^{\beta\tau^\gamma} dm \leq C_\tau$ , then  $\int e^{\beta'h_\mathcal{A}^\gamma} d\mathbb{P}_\mathcal{A} \leq C$ , where the constants  $\beta' \in (0, \beta]$  and  $C > 0$  depend continuously on  $C_\tau$ ,  $\beta$ ,  $\gamma$ ,  $\lambda$ ,  $K$  and  $\eta$ .

*Proof.* Such a decomposition is constructed in [10, Lemma 4.3, Propositions 4.4 and 4.5]. It is implicit in [10] that  $m_w$  is supported on  $Y_w$ .

The bound for the strong polynomial moment of  $h_\mathcal{A}$  is not proved in [10], but the proof is identical to the one for the weak polynomial moments [10, Proposition 4.4].  $\square$

To prove the exactly exponential moments case in Theorem 3.4, we obtain a special version of Lemma 4.2:

**Lemma 4.3.** *Suppose that  $\int e^{\beta\tau} dm < \infty$  with some  $\beta > 0$ . Then the measure  $\mathbb{P}_\mathcal{A}$  in Lemma 4.2 can be chosen so that*

$$\mathbb{P}_\mathcal{A}(h_\mathcal{A} = \ell) = \begin{cases} \theta^{-1}(1 - \theta)\theta^{\ell/N}, & \ell \in \mathbb{N} \\ 0, & \text{else} \end{cases}$$

with some  $N \in \mathbb{N}$  and  $0 < \theta < 1$ .

Our proof of Lemma 4.3 uses a rather delicate technical adaptation of the argument in [11, Section 4]. It is carried out in Appendix A.

Let  $\mathbb{P}_\mathcal{A}$  and  $\{m_w\}$  be as in Lemmas 4.2 or 4.3. Let  $\xi = \lambda^{-1}$ . According to Remark 3.2,  $\mathcal{A}$ ,  $\mathbb{P}_\mathcal{A}$ ,  $h_\mathcal{A}$  and  $\xi$  define a Bernoulli Young tower  $f: \Delta \rightarrow \Delta$ . To prove Theorem 3.4, it remains to construct the semiconjugacy  $\pi: \Delta \rightarrow \Lambda$ .

**4.2. Semi-conjugacy.** For words  $w_0, \dots, w_n \in \mathcal{A}$ , let  $w_0 \cdots w_n$  denote their concatenation. Then  $|w_0 \cdots w_n| = |w_0| + \dots + |w_n|$  and  $h_\mathcal{A}(w_0 \cdots w_n) = h_\mathcal{A}(w_0) + \dots + h_\mathcal{A}(w_n)$ .

For  $x = (w_0, w_1, \dots) \in X$ , let  $\pi_X(x)$  be the only point in  $\bigcap_{n \geq 0} Y_{w_0 \cdots w_n}$ . Note that  $\{Y_{w_0 \cdots w_n}\}_{n \geq 0}$  is a nested sequence of closed sets in the complete metric space  $Y$  and  $\text{diam } Y_{w_0 \cdots w_n} \rightarrow 0$ , thus  $\pi_X(x)$  is well defined.

Define  $\pi: \Delta \rightarrow \Lambda$  by

$$(3) \quad \pi((w_0, w_1, \dots), \ell) = T_{w_0}^\ell(\pi_X(w_0, w_1, \dots)).$$

Note that  $\pi_X: \Delta_0 \rightarrow Y$  is a restriction of  $\pi$ .

**Proposition 4.4.**  *$\pi$  is Lipschitz: for all  $a, b \in \Delta$ ,*

$$d_\Lambda(\pi(a), \pi(b)) \leq C_\Lambda d(a, b),$$

where  $C_\Lambda = \lambda C_\ell \text{diam } \Lambda$ .

*Proof.* Let  $a = (x_1, j)$  and  $b = (x_2, k)$ , where

$$x_1 = (w_{1,0}, w_{1,1}, \dots) \quad \text{and} \quad x_2 = (w_{2,0}, w_{2,1}, \dots).$$

If  $j \neq k$  or  $w_{1,0} \neq w_{2,0}$ , then  $d(a, b) = 1$  and the statement is trivial.

Suppose now that  $j = k$  and  $w_{1,0} = w_{2,0}$ . Let  $n = s(x_1, x_2)$ . Note that  $n \geq 1$  and

$$j = k < h(x_1) = h(x_2) = h_\mathcal{A}(w_{1,0}) = h_\mathcal{A}(w_{2,0}).$$

Observe that  $\pi_X(x_i) \in Y_{w_{1,0} \cdots w_{1,n-1}}$  and  $T_{Y, w_{1,0}}(\pi_X(x_i)) \in Y_{w_{1,1} \cdots w_{1,n-1}}$  for  $i = 1, 2$ . Also,  $\text{diam } Y_{w_{1,1} \cdots w_{1,n-1}} \leq \lambda^{-(n-1)} \text{diam } Y$ . Then

$$\begin{aligned} d_\Lambda(\pi(a), \pi(b)) &= d_\Lambda(T_{w_{1,0}}^j(\pi_X(x_1)), T_{w_{1,0}}^j(\pi_X(x_2))) \leq C_\ell \text{diam } Y_{w_{1,1} \cdots w_{1,n-1}} \\ &\leq C_\ell \lambda^{-(n-1)} \text{diam } Y = \lambda C_\ell \text{diam } Y d(a, b). \end{aligned}$$

$\square$

**Proposition 4.5.**  $(\pi_X)_* \mathbb{P}_X = m$ .

*Proof.* Recall that we have the disintegration  $m = \sum_{w \in \mathcal{A}} \mathbb{P}_\mathcal{A}(w) m_w$ .

Let  $w_0 \in \mathcal{A}$ . Since  $T_{Y, w_0}: Y_{w_0} \rightarrow Y$  is a bijection and  $(T_{Y, w_0})_* m_{w_0} = m$ , we can write  $m_{w_0} = \sum_{w_1 \in \mathcal{A}} \mathbb{P}_\mathcal{A}(w_1) m_{w_0, w_1}$ , where  $m_{w_0, w_1}$  are probability measures supported on  $Y_{w_0 w_1}$  such that  $(T_{Y, w_1} \circ T_{Y, w_0})_* m_{w_0, w_1} = m$ . Continuing with  $m_{w_0, w_1}$  and further recursively, we obtain for each  $n \geq 1$  a disintegration

$$m = \sum_{w_0, \dots, w_n \in \mathcal{A}} \mathbb{P}_\mathcal{A}(w_0) \cdots \mathbb{P}_\mathcal{A}(w_n) m_{w_0, \dots, w_n},$$

where  $m_{w_0, \dots, w_n}$  are probability measures supported on  $Y_{w_0 \dots w_n}$  such that

$$(T_{Y, w_n} \circ \cdots \circ T_{Y, w_0})_* m_{w_0, \dots, w_n} = m.$$

Let  $w \in \mathcal{A}$  and  $n = |w|$ . Then either  $Y_{w_0 \dots w_n} \subset Y_w$ , or  $m(Y_{w_0 \dots w_n} \cap Y_w) = 0$  for all  $w_0, \dots, w_n \in \mathcal{A}$ . Thus

$$m(Y_w) = \sum_{\substack{w_0, \dots, w_n \in \mathcal{A}: \\ Y_{w_0 \dots w_n} \subset Y_w}} \mathbb{P}_\mathcal{A}(w_0) \cdots \mathbb{P}_\mathcal{A}(w_n) = \mathbb{P}_X(\pi_X^{-1}(Y_w)).$$

Thus  $(\pi_X)_* \mathbb{P}_X$  agrees with  $m$  on all sets in  $\mathcal{Y} = \{Y_w : w \in \mathcal{A}\}$ . It is easy to see that, being countable,  $\mathcal{Y}$  generates the Borel sigma algebra on  $Y$ . Also,  $(\pi_X)_* \mathbb{P}_X$  is a pre-measure on the algebra generated by  $\mathcal{Y}$ . By Carathéodory's extension theorem,  $(\pi_X)_* \mathbb{P}_X = m$ .  $\square$

**Proposition 4.6.**  $\mathbb{P}$ -almost surely,  $T \circ \pi = \pi \circ f$ .

*Proof.* Suppose that  $a = (x, \ell) \in \Delta$ , and  $x = (w_0, w_1, \dots)$ . By Proposition 4.5,  $T_{w_0}(\pi(a)) = T(\pi(a))$  for  $\mathbb{P}$ -almost every  $a$ . If  $\ell < h(x) - 1$ , then  $f(a) = (x, \ell + 1)$ , and

$$\pi(f(a)) = T_{w_0}^{\ell+1}(\pi_X(x)) = T_{w_0}(\pi(a)).$$

If  $\ell = h(x) - 1$ , then also

$$\pi(f(a)) = \pi_X(f_X(x)) = T_{Y, w_0}(\pi_X(x)) = T_{w_0}^{\ell+1}(\pi_X(x)) = T_{w_0}(\pi(a)).$$

Thus  $\pi(f(a)) = T(\pi(a))$  for  $\mathbb{P}$ -almost every  $a$ .  $\square$

**Proposition 4.7.**  $\pi_* \mathbb{P} = \nu$ .

*Proof.* We use the fact that  $\nu$  is the unique  $T$ -invariant ergodic probability measure on  $\Lambda$ , with respect to which  $m$  is absolutely continuous.

Since  $\mathbb{P}$  is  $f$ -invariant and ergodic, it follows from Proposition 4.6 that  $\pi_* \mathbb{P}$  is  $T$ -invariant and ergodic. Since  $\mathbb{P}_X$  is absolutely continuous with respect to  $\mathbb{P}$  and  $\pi_* \mathbb{P}_X = m$ , using Proposition 4.5 we obtain that  $m$  is absolutely continuous with respect to  $\pi_* \mathbb{P}$ . Thus  $\pi_* \mathbb{P} = \nu$ .  $\square$

## 5. PROOF OF THEOREM 2.1

By Theorem 3.4, it is enough to prove the ASIP for Hölder continuous observables on a Bernoulli Young tower with

$$\mathbb{P}_\mathcal{A}(h_\mathcal{A} = n) = \begin{cases} \theta(1 - \theta)^{-1}(1 - \theta)^{n/N}, & n \in \{N, 2N, 3N, \dots\} \\ 0, & \text{else} \end{cases}$$

with  $N \in \mathbb{N}$  and  $0 < \theta < 1$ . Then  $T$  is  $N$ -periodic, and without loss we assume that  $N = 1$ .

Let  $(\Omega, \mathbb{P}_\Omega)$  be a probability space supporting random variables  $A_n: \Omega \rightarrow \mathcal{A}$ ,  $n \geq 1$ , such that for  $a \in \mathcal{A}$ ,

$$\mathbb{P}(A_n = a) = \begin{cases} 0, & h(a) \neq n, \\ \frac{\mathbb{P}_{\mathcal{A}}(a)}{\mathbb{P}_{\mathcal{A}}(\cup\{a \in \mathcal{A}: h_{\mathcal{A}}(a) = n\})}, & \tau(a) = n. \end{cases}$$

That is,  $A_n$  is a random element of  $\mathcal{A}$  chosen among those with  $h_{\mathcal{A}} = n$  with respect to the appropriately conditioned measure  $\mathbb{P}_{\mathcal{A}}$ .

Let  $Z = \{0, 1\}$  and  $\mathbb{P}_Z$  be the probability measure on  $Z$  given by  $\mathbb{P}_Z(0) = 1 - \theta$  and  $\mathbb{P}_Z(1) = \theta$ . Let  $D = \Omega^{\mathbb{Z}} \times Z^{\mathbb{Z}}$  with the product probability measure  $\mathbb{P}_D = \mathbb{P}_\Omega^{\mathbb{Z}} \times \mathbb{P}_Z^{\mathbb{Z}}$ , and let  $\sigma: D \rightarrow D$  be the left shift.

For  $x = ((\dots, \omega_{-1}, \omega_0, \omega_1, \dots), (\dots, z_{-1}, z_0, z_1, \dots)) \in D$  let

$$t_0(x) = \sup\{k \leq 0: z_k = 1\} \quad \text{and} \quad t_n(x) = \inf\{k > t_{n-1}: z_k = 1\}, \quad n \geq 1.$$

Then  $t_n, n \geq 0$  are finite  $\mathbb{P}_D$ -almost surely.

Define  $g: D \rightarrow \Delta$  by  $g(x) = (y, -t_0(x))$ , where  $y \in X$ ,

$$y = (A_{t_1(x)-t_0(x)}(\omega_{t_0(x)}), A_{t_2(x)-t_1(x)}(\omega_{t_1(x)}), \dots)$$

Observe that  $g$  is a measure preserving semiconjugacy between  $\sigma: D \rightarrow D$  and  $T: \Delta \rightarrow \Delta$ , with measures  $\mathbb{P}_D$  on  $D$  and  $\mathbb{P}$  on  $\Delta$ .

Suppose that  $v: \Delta \rightarrow \mathbb{R}$  is a centered Hölder observable. Let  $\tilde{v} = v \circ g$  and  $\tilde{v}_n = \sum_{k=0}^{n-1} \tilde{v} \circ \sigma^k$ . We prove the ASIP for  $\tilde{v}_n$  using [1].

For  $x = ((\dots, \omega_{-1}, \omega_0, \omega_1, \dots), (\dots, z_{-1}, z_0, z_1, \dots)) \in D$  and  $k \in \mathbb{Z}$  let

$$D_k(x) = \{x' = ((\dots, \omega'_{-1}, \omega'_0, \omega'_1, \dots), (\dots, z'_{-1}, z'_0, z'_1, \dots)) \in D: \\ \omega'_j = \omega_j \text{ and } z'_j = z_j \text{ for all } j \neq k\}.$$

Fix  $p > 2$  and let

$$\delta_{k,p} = \int_D \sup_{x' \in D_k(x)} d^p(g(x), g(x')) d\mathbb{P}(x).$$

**Proposition 5.1.** *There exists  $C_\delta > 0$  and  $0 < \theta_\delta < 1$  such that  $\delta_{k,p} \leq C_\delta \theta_\delta^{|k|}$  for all  $k$ .*

*Proof.* Suppose that  $x \in D$  and  $x' \in D_k(x)$ , and

$$x = ((\dots, \omega_{-1}, \omega_0, \omega_1, \dots), (\dots, z_{-1}, z_0, z_1, \dots)).$$

Let

$$g(x) = ((w_0, w_1, \dots), \ell), \\ g(x') = ((w'_0, w'_1, \dots), \ell').$$

Recall that  $\text{diam } \Delta = 1$ .

Suppose first that  $k \geq 1$ . Let  $c_k = \sum_{j=1}^{k-1} z_k$ . Then  $w_j = w'_j$  for all  $0 \leq j \leq c_k - 1$ . If  $c_k \geq 1$ , then  $\ell = \ell'$  and  $d(g(x), g(x')) \leq \xi^{c_k}$ . If  $c_k = 0$ , then  $d(g(x), g(x')) \leq 1$ , so  $d(g(x), g(x')) \leq \xi^{c_k}$  for all  $c_k$ .

Next,  $z_j$  can be treated as independent identically distributed random variables, so

$$\int_D \xi^{pc_k} d\mathbb{P}_D(x) = \left[ \int_D \xi^{pz_1} d\mathbb{P}_D(x) \right]^{k-1} = (1 - \theta + \theta \xi^p)^{k-1}.$$

The result for  $k \geq 1$  follows.

Suppose now that  $k \leq 0$ . Then  $g(x) \neq g(x')$  only when  $t_0(x) \leq k$ . Next,

$$\mathbb{P}_D(t_0(x) \leq k) = \mathbb{P}_D(z_0 = z_{-1} = \dots = z_{k-1} = 0) = (1 - \theta)^k.$$

The result for  $k \leq 0$  follows. □

Proposition 5.1 verifies that, in notation of [1],  $\delta_{i,p}$  and  $\Theta_{i,p}$  are exponentially small in  $|i|$ . By [1, Theorem 2.1] (using [1, Corollary 2.1] to simplify verification of the assumptions),  $\tilde{v}_n$ , defined on the probability space  $(D, \mathbb{P}_D)$  satisfies the ASIP with rate  $o(n^{1/p})$ . Then so does  $v_n$  on  $(\Delta, \mathbb{P})$ .

The proof of Theorem 2.1 is complete.

## APPENDIX A. PROOF OF LEMMA 4.3

Our argument is based on [11, Section 4], and here we work in their notations, which are different from the rest of this paper.

In this section,  $T: \Lambda \rightarrow \Lambda$  is a nonuniformly expanding map as in Section 2,  $F: Y \rightarrow Y$ ,  $F = T_Y$  is the induced map, and  $f: \Delta \rightarrow \Delta$  is the Young tower,

$$\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \leq \ell < \tau(y)\},$$

$$f(y, \ell) = \begin{cases} (y, \ell + 1), & \ell < \tau(y) - 1, \\ (Fy, 0), & \ell = \tau(y) - 1. \end{cases}$$

Let  $\bar{\tau} = \int_Y \tau dm$ . Let  $m_\Delta$  be the probability measure on  $\Delta$  given by  $m_\Delta(A \times \{\ell\}) = \bar{\tau}^{-1}m(A)$  for all  $\ell \geq 0$  and  $A \subset \{y \in Y : \tau(y) \geq \ell + 1\}$ .

Let  $L: L^1(m_\Delta) \rightarrow L^1(m_\Delta)$  be the transfer operator corresponding to  $f$  and  $m_\Delta$ , so  $\int L\phi\psi dm_\Delta = \int \phi\psi \circ f dm$  for all  $\phi \in L^1$  and  $\psi \in L^\infty$ .

Without loss we assume that  $f$  is mixing (otherwise we switch to a power of  $f$  which is mixing).

Fix constants  $R > 0$  and  $\xi \in (0, e^{-R})$ , such that  $R(1 - \xi e^R) \geq K + \lambda^{-1}R$ . Let  $\Delta_\ell = \{(y, k) \in \Delta : k = \ell\}$ . For  $\psi: \Delta \rightarrow [0, \infty)$ , define

$$|\psi|_{\eta, \ell} = \sup_{n \geq 0} \sup_{(y, n) \neq (y', n) \in \Delta_n} \frac{|\log \psi(y, n) - \log \psi(y', n)|}{d(y, y')^\eta},$$

where  $\log 0 = -\infty$  and  $\log 0 - \log 0 = 0$ . Note that for a countable collection  $\psi_k$  of nonnegative functions,  $|\sum_k \psi_k|_{\eta, \ell} \leq \max_k |\psi_k|_{\eta, \ell}$ .

For  $a \in \alpha$ , let  $S_a = \{(y, k) \in \Delta : y \in a \text{ and } k = \tau(y) - 1\}$ , and let  $\varkappa$  be a partition of  $\Delta$  generated by  $\{S_a\}, a \in \alpha$  and  $\{\Delta_\ell\}, \ell \geq 0$ . Let also  $\varkappa^n = \bigvee_{k=0}^{n-1} f^{-k} \varkappa$ . Then  $\varkappa_0$  is a trivial partition, and for every  $n \geq 1$  and  $a \in \varkappa_n$ , there exists  $\ell \geq 0$  such that  $f^n: a \rightarrow \Delta_\ell$  is a bijection.

**Proposition A.1.** *Suppose that  $\psi: \Delta \rightarrow [0, \infty)$  with  $|\psi|_{\eta, \ell} \leq R$ . Let  $n \geq 1$ ,  $a \in \varkappa^n$  and  $\psi_a = \psi 1_a$ . Then*

- (a)  $e^{-R\bar{\tau}} \int_{\Delta_0} \psi dm_\Delta \leq \psi 1_{\Delta_0} \leq e^{R\bar{\tau}} \int_{\Delta_0} \psi dm_\Delta$ .
- (b)  $|L^n \psi_a|_{\eta, \ell} \leq R$ .
- (c) If  $t \in [0, \xi]$ , then  $\psi'_a = L^n \psi_a - t \bar{\tau} \int_{\Delta_0} L^n \psi_a dm_\Delta 1_{\Delta_0}$  is nonnegative and  $|\psi'_a|_{\eta, \ell} \leq R$ .

*Proof.* This is a minor modification of [11, Proposition 4.1].  $\square$

Let  $\mathcal{A}$  be the set of observables  $\psi: \Delta \rightarrow [0, \infty)$  such that  $|\psi|_\infty \leq e^{R\bar{\tau}} \int_{\Delta} \psi dm_\Delta$  and  $|\psi|_{\eta, \ell} \leq R$ .

For  $n \geq 0$ , let  $\mathcal{A}^n$  denote the set of observables  $\psi: \Delta \rightarrow [0, \infty)$  such that  $L^n \psi \in \mathcal{A}$  and  $|L^n(\psi 1_a)|_{\eta, \ell} \leq R$  for every  $a \in \varkappa^n$ .

**Corollary A.2.**

- (a) If  $\psi: \Delta \rightarrow [0, \infty)$  is supported on  $\Delta_0$  and  $|\psi|_{\eta, \ell} \leq R$ , then  $\psi \in \mathcal{A}$ .
- (b) If  $\psi \in \mathcal{A}$ , then  $L\psi \in \mathcal{A}$ .
- (c) If  $\psi \in \mathcal{A}^n$ , then  $\psi \in \mathcal{A}^k$  for all  $k \geq n$ .



(d) If  $\psi, \psi' \in \mathcal{A}^n$  and  $t \geq 0$ , then  $\psi + \psi'$  and  $t\psi$  belong in  $\mathcal{A}^n$ .

*Proof.* See [11, Corollary 4.2].  $\square$

**Lemma A.3.** *There exist  $N \geq 1$  and  $\epsilon > 0$  such that*

- (a)  $\int_{\Delta_0} \psi dm_\Delta \geq \epsilon \int_\Delta \psi dm_\Delta$  for all  $\psi \in L^N \mathcal{A}$ ,
- (b)  $(1 - \xi\epsilon) \left(\frac{1-\xi}{1-\xi\epsilon}\right)^n \geq e^{R\bar{\tau}} m_\Delta(\cup_{\ell=Nn}^\infty \Delta_\ell)$  for all  $n \geq 1$ .

*Proof.* (a) is proved in [11, Lemma 4.5]. Following the proof, we are free to choose  $\epsilon$  as small as needed and  $N$  as large as needed. By assumptions of Lemma 4.3,  $m_\Delta(\cup_{\ell=n}^\infty \Delta_\ell)$  decays exponentially in  $n$ , thus we can choose  $N$  and  $\epsilon$  so that (b) is satisfied.  $\square$

Further we assume that  $N$  and  $\epsilon$  are as in Lemma A.3. Define  $\mathcal{B} = L^N \mathcal{A}$ . Note that  $L\mathcal{B} \subset \mathcal{B}$ . For  $n \geq 0$  let  $\mathcal{B}^n$  denote the set of observables  $\psi: \Delta \rightarrow [0, \infty)$  such that  $L^n \psi \in \mathcal{B}$  and  $|L^n(\psi 1_a)|_{\eta, \ell} \leq R$  for every  $a \in \mathcal{X}^n$ .

*Remark A.4.* If  $\psi \in \mathcal{B}$ , then  $L\psi \in \mathcal{B}$ . If  $\psi \in \mathcal{B}^n$ , then  $\psi \in \mathcal{B}^k$  for all  $k \geq n$ .

Define a sequence  $p_n, n \geq -1$  by

$$p_{-1} = \xi\epsilon \quad \text{and} \quad p_n = \begin{cases} (1 - \xi)\epsilon \left(\frac{1-\epsilon}{1-\xi\epsilon}\right)^{n/N}, & n \in N\mathbb{Z} \\ 0, & n \notin N\mathbb{Z} \end{cases} \quad \text{for } n \geq 0.$$

Let  $t_n = 1 - \sum_{k=-1}^{n-1} p_k$  for  $n \geq 1$ . Then  $\sum_{k=-1}^\infty p_k = 1$ ,  $t_1 = 1 - \epsilon$  and for  $n \geq 2$  using Lemma A.3 we obtain

$$(4) \quad t_n \geq 1 - \sum_{k=-1}^{Nn-1} p_k = (1 - \xi\epsilon) \left(\frac{1-\epsilon}{1-\xi\epsilon}\right)^n \geq \min\{t_1, e^{R\bar{\tau}} m_\Delta(\cup_{\ell=n}^\infty \Delta_\ell)\}.$$

Let  $E_0 = \Delta_0$  and  $E_k = \{(y, \ell) \in \Delta : \ell = \tau(y) - k, \ell \geq 1\}$  for  $k \geq 1$ . Then  $\{E_0, E_1, \dots\}$  defines a partition of  $\Delta$  and  $m_\Delta(E_k) = m_\Delta(\Delta_k)$  for all  $k$ .

**Proposition A.5.** *If  $\psi \in \mathcal{B}$  with  $\int_\Delta \psi dm_\Delta = 1$ , then  $\int_{\cup_{\ell=n}^\infty E_\ell} \psi dm_\Delta \leq t_n$ , for  $n \geq 1$ .*

*Proof.* See [11, Proposition 4.6].  $\square$

**Proposition A.6.** *Let  $p_j, q_j \in [0, \infty)$  be sequences such that  $\sum_{j=0}^\infty p_j = \sum_{j=0}^\infty q_j < \infty$  and  $\sum_{j=0}^k q_j \geq \sum_{j=0}^k p_j$  for all  $k \geq 0$ . Then there exist  $s_{k,j} \in [0, 1]$ ,  $0 \leq j \leq k$ , such that  $\sum_{j=0}^k s_{k,j} q_j = p_k$  for all  $k \geq 0$  and  $\sum_{k=j}^\infty s_{k,j} = 1$  for all  $j \geq 0$ .*

*Proof.* See [11, Proposition 4.7].  $\square$

**Lemma A.7.** *Let  $\psi \in \mathcal{B}^n$  for some  $n \geq 0$ . Then  $\psi = \sum_{k=-1}^\infty \psi_k$ , where  $\psi_k: \Delta \rightarrow [0, \infty)$  are such that*

- (a)  $L^n(\psi_{-1} 1_a) = c_a 1_{\Delta_0}$  for all  $a \in \mathcal{X}^n$ , where  $c_a$  are nonnegative constants,
- (b)  $\sum_{a \in \mathcal{X}^n} c_a = p_{-1} \bar{\tau} \int_\Delta \psi dm_\Delta$ ,
- (c)  $\psi_k \in \mathcal{A}^{n+k}$  for all  $k \geq 0$ ,
- (d)  $\int_\Delta \psi_k dm_\Delta = p_k \int_\Delta \psi dm_\Delta$  for all  $k \geq -1$

*Proof.* We follow the proof of [11, Lemma 4.8]. Suppose without loss that  $\int_\Delta \psi dm_\Delta = 1$ . Define

$$t = p_{-1} / \int_{\Delta_0} L^n \psi dm_\Delta = \xi\epsilon / \int_{\Delta_0} L^n \psi dm_\Delta.$$

By Lemma A.3,  $\int_{\Delta_0} L^n \psi dm_\Delta \geq \epsilon$ , so  $t \in [0, \xi]$ .

Under convention that  $0/0 = 0$ , let

$$\psi_{-1} = t\bar{\tau} \sum_{a \in \mathcal{X}^n} \left( \frac{\int_{\Delta_0} L^n(\psi 1_a) dm_{\Delta}}{L^n(\psi 1_a)} \circ f^n \right) \psi 1_a.$$

Then properties (a) and (b) are satisfied.

Let  $g = \psi - \psi_{-1}$  and  $g_k = g 1_{T^{-n}E_k}$  for  $k \geq 0$ . Then  $L^{n+k}g_k$  is supported on  $\mathcal{A}$  and  $|L^{n+k}(g_k 1_a)|_{\eta, \ell} \leq R$  for every  $a \in \mathcal{X}^n$ . By Corollary A.2,  $g_k \in \mathcal{A}^{n+k}$ .

Let  $q_k = \int_{\Delta} g_k dm_{\Delta}$ . Then  $\sum_{k=0}^{\infty} q_k = \sum_{k=0}^{\infty} p_k$  and by Proposition A.5,  $\sum_{k=0}^n q_k \geq \sum_{k=0}^n p_k$  for all  $n \geq 0$ . Choose  $s_{k,j} \in [0, 1]$  as in Proposition A.6, and define  $\psi_k : \Delta \rightarrow [0, \infty)$ ,  $k \geq 0$ , by

$$\psi_k = \sum_{j=0}^k s_{k,j} g_j.$$

Then (d) holds for all  $k$ . Corollary A.2 implies (c).  $\square$

Let  $\mathbb{W}$  be the countable set of all finite words in the alphabet  $\mathbb{N}_0$  including the zero length word, and let  $\mathbb{W}_k$  be the subset consisting of words of length  $k$ . Let  $\mathbb{P}_{\mathbb{W}}$  be the probability measure on  $\mathbb{W}$  given for  $w = w_1 \cdots w_k \in \mathbb{W}_k$  by  $\mathbb{P}_{\mathbb{W}}(w) = p_{-1} p_{w_1} \cdots p_{w_k}$ . Define  $r : \mathbb{W} \rightarrow \mathbb{N}_0$  by  $r(w) = \Sigma w + N|w|$ , where  $\Sigma w = w_1 + \cdots + w_k$  and  $|w| = k$  for  $w = w_1 \cdots w_k$ .

**Proposition A.8.** *Let  $\psi \in \mathcal{B}$  with  $\int_{\Delta} \psi dm_{\Delta} = 1$ . Then  $\psi = \sum_{w \in \mathbb{W}} \psi_w$ , where  $\psi_w : \Delta \rightarrow [0, \infty)$  are such that*

- (a)  $\int_{\Delta} \psi_w dm_{\Delta} = \mathbb{P}_{\mathbb{W}}(w)$ ,
- (b)  $L^{r(w)} \psi_w = \mathbb{P}_{\mathbb{W}}(w) \bar{\tau} 1_{\Delta_0}$ ,
- (c)  $L^{r(w)}(\psi_w 1_a) = c_{w,a} 1_{\Delta_0}$  for all  $a \in \mathcal{X}^{r(w)}$ , where  $c_{w,a}$  are nonnegative constants.

*Proof.* Proof is identical to [11, Proposition 4.9] except for condition (c), which is guaranteed by Lemma A.7.  $\square$

**Definition A.9.** We say that a random variable  $X$  has geometric distribution with parameter  $\theta \in (0, 1)$  (or  $X \sim \text{Geom}(\theta)$ ), if  $X$  takes values in  $\mathbb{N}_0$  and  $\mathbb{P}(X = n) = (1 - \theta)^n \theta$  for  $n \geq 0$ .

**Proposition A.10.** *Suppose that  $Y = \sum_{k=1}^M (1 + X_k)$ , where  $M \sim \text{Geom}(\theta_M)$  and  $X_k \sim \text{Geom}(\theta_X)$  are independent random variables. Let  $\eta_2 = \theta_X \theta_M$  and  $\eta_1 = \frac{\theta_M - \eta_2}{1 - \eta_2}$ . Then*

$$\mathbb{P}(Y = n) = \begin{cases} \eta_1 + (1 - \eta_1)\eta_2, & n = 0 \\ (1 - \eta_1)\eta_2(1 - \eta_2)^n, & n \geq 1 \end{cases}.$$

*Proof.* We compute the probability generating function of  $Y$ . For  $z \in \mathbb{R}$ ,

$$\mathbb{E}(z^Y) = \mathbb{P}(R = 0) + \mathbb{P}(R \geq 1) \mathbb{E}(z^{1+X_1}) \mathbb{E}(z^Y).$$

Using that

$$\mathbb{E}(z^{1+X_1}) = \sum_{k=0}^{\infty} \mathbb{P}(X_1 = k) z^{k+1} = \frac{\theta_X z}{1 - (1 - \theta_X)z},$$

we obtain

$$\mathbb{E}(z^Y) = \eta_1 + (1 - \eta_1) \frac{\eta_2}{1 - (1 - \eta_2)z}.$$

Now,  $\mathbb{P}(Y = n)$  is the coefficient at  $z^n$  in the above expression.  $\square$

**Proposition A.11.** *There exist constants  $0 < \theta < 1$  and  $C_1, C_2 > 0$  such that*

$$\mathbb{P}(r = n) = \begin{cases} C_1, & n = 0 \\ C_2 \theta^{n/N}, & n \in \mathbb{N} \\ 0, & \text{else} \end{cases}$$

*Proof.* Recall that  $\mathbb{W}_k$  is the subset of  $\mathbb{W}$  consisting of words of length  $k$ . Then  $\mathbb{P}_{\mathbb{W}}(\mathbb{W}_k) = (1 - p_{-1})^k p_{-1}$ . Elements of  $\mathbb{W}_k$  have the form  $w_1 \cdots w_k$  where  $w_1, \dots, w_k$  can be regarded as independent identically distributed random variables, drawn from  $\mathbb{N}_0$  with distribution

$$\mathbb{P}(w_1 = n) = p_n / (1 - p_{-1}) = \begin{cases} \theta_1 (1 - \theta_1)^{n/N}, & n \in N\mathbb{N}_0 \\ 0, & \text{else} \end{cases},$$

where  $\theta_1 = \frac{(1-\xi)\epsilon}{1-\xi\epsilon}$ . In other words,  $w_1/N \sim \text{Geom}(\theta_1)$ .

Then the random variable  $r/N$  on  $\mathbb{W}$  has the same distribution as  $Y$  in Proposition A.10 with  $\theta_M = p_{-1}$  and  $\theta_X = \theta_1$ . The result follows.  $\square$

We are ready to complete the proof of Lemma 4.3. Let  $\psi = dm/dm_{\Delta} = \bar{\tau}1_{\Delta_0}$  and  $\psi = \sum_{w \in \mathbb{W}} \psi_w$  be the decomposition from Proposition A.8.

Then  $\psi = \sum_{w \in \mathbb{W}} \sum_{a \in A(w)} \psi_w 1_a$ , where  $A(w) = \{a \in \mathcal{A}^{r(w)} : a \subset \Delta_0 \text{ and } f^{r(w)}a = \Delta_0\}$ . To every  $w \in \mathbb{W}$  and  $a \in A(w)$  there corresponds  $u \in \mathcal{A}$  such that  $a = Y_u$  (modulo zero  $m$  measure) and  $r(w) = h_{\mathcal{A}}(u)$ . Thus we can write

$$m = \sum_{u \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(u) m_u,$$

where  $m_u$  are probability measures supported on  $Y_u$  and  $\mathbb{P}_{\mathcal{A}}$  is a probability measure on  $\mathcal{A}$  such that  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} = n) = \mathbb{P}_{\mathbb{W}}(r = n)$  for all  $n$ .

The result of Lemma 4.3 follows from Proposition A.11.

#### ACKNOWLEDGEMENTS

This research was supported in part by a European Advanced Grant *StochExtHomog* (ERC AdG 320977). The author is grateful to Ian Melbourne for helpful discussions.

#### REFERENCES

- [1] I. Berkes, W. Liu, W. Wu, *Komlós-Major-Tusnády approximation under dependence*, Ann. Probab. **42** (2014), 794–817.
- [2] N. Chernov, C. Haskell, *Nonuniformly hyperbolic K-systems are Bernoulli*, Ergodic Theory Dynam. Systems **16** (1996), 19–44.
- [3] J.-P. Conze, S. Le Borgne, *Méthode de martingales et flow géodésique sur une surface de courbure constante négative*, Ergodic Theory Dynam. Systems **21** (2001), 421–441.
- [4] C. Cuny, F. Merlevède, *Strong invariance principles with rate for “reverse” martingales and applications*, J. Theor. Probab. (2015), 137–183.
- [5] M. Denker, W. Philipp, *Approximation by Brownian motion for Gibbs measures and flows under a function*, Ergodic Theory Dynam. Systems **4** (1984), 541–552.
- [6] M. J. Field, I. Melbourne, A. Török, *Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions*, Ergodic Theory Dynam. Systems **23** (2003), 87–110.
- [7] S. Gouëzel, *Almost sure invariance principle for dynamical systems by spectral methods*, Ann. Probab. **38** (2010), 1639–1671.
- [8] F. Hofbauer, G. Keller, *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z. **180** (1982), 119–140.
- [9] J. Komlós, P. Major, G. Tusnády, *An approximation of partial sums of independent RV’s and the sample DF. I; II*, Z. Wahrscheinlichkeitstheor. verw. Geb. **32** (1975), 111–131; **34** (1976), 34–58.
- [10] A. Korepanov, *Equidistribution for nonuniformly expanding systems*, Preprint, 2017. arXiv:1701.03652.
- [11] A. Korepanov, Z. Kosloff, I. Melbourne, *Explicit coupling argument for nonuniformly hyperbolic transformations*, Proc. Roy. Soc. Edinburgh Sect. A., to appear. arXiv:1602.03795.
- [12] A. Korepanov, Z. Kosloff, I. Melbourne, *Martingale-coboundary decomposition for families of dynamical systems*, Preprint, 2016. arXiv:1608.01853.

- [13] I. Melbourne, M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Commun. Math. Phys. **260** (2005), 131–146.
- [14] I. Melbourne, M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Ann. Probab. **37** (2009), 478–505.
- [15] W. Philipp, W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Amer. Math. Soc. Mem. **161** (1975).
- [16] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. **147** (1998), 585–650.
- [17] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188.

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UK  
*E-mail address:* `a.korepanov@warwick.ac.uk`